NOVEL ACETYLCHOLINESTERASE GENE RESPONSIBLE FOR INSECTICIDE RESISTANCE AND APPLICATIONS THEREOF

The present invention relates to a novel acetylcholinesterase gene responsible for insecticide resistance, in particular in mosquitoes, to the products of this gene (cDNA, protein) and to the applications thereof, in particular for screening novel insecticides and for the genetic detection of resistance to organophosphorus compounds and/or to carbamates in mosquito populations.

Acetylcholinesterase (AChE, E.C. 3.1.1.7) is an essential enzyme which hydrolyzes acetylcholine in the synapses, thus terminating cholinergic transmissions at neuronal or neuromuscular junctions. The inhibition of AChE prevents the deactivation of the synaptic signal, thus resulting in a loss of control of cholinergic transmission. The biology of acetylcholinesterase has been greatly studied in invertebrates, and in particular insects, since this enzyme is the target for the main classes of pesticides used, organophosphorus compounds and carbamates. However, the massive use of pesticides over the past decades has caused resistant species to emerge. Among the mechanisms of resistance, the selection of mutations making AChE insensitive to insecticides has been observed in many cases (for a review, see Fournier et al., Comp. Biochem. Physiol., 1994, 108, 19-31).

In order to precisely determine the nature of the AChE that is a target for insecticides, and also the mutations responsible for the resistance to the latter, the genes encoding AChEs (*ace* genes) have been isolated in various arthropod (insect and arachnid) species.

The first *ace* gene was identified in drosophila (*Drosophila melanogaster*), by reverse genetics (Hall et al., EMBO J., 1986, 5, 2949-2954). The proof that this gene was involved in insecticide

resistance was provided by the demonstration of amino acid substitutions in the AChE of resistant drosophila, conferring insensitivity to cholinergic insecticides (Mutéro et al., P.N.A.S., 1994, 91, 5922-5926). The studies in *D. melanogaster* therefore appear to indicate the presence of a single *ace* gene in insects, encoding the AChE that is a target for cholinergic insecticides.

However, with the exception of the *ace* gene of two other insects, *Musca domestica* (Williamson et al., 1992, in *Multidisciplinary approaches to cholinesterase functions*, Eds Schafferman A. & Velan B., Plenum Press, New-York, pp 83-86; Walsh et al., Biochem. J., 2001, 359, 175-181; Kozaki et al., Insect Biochem. Mol. Biol., 2001, 31, 991-997) and *Bactrocera oleae* (Vontas et al., Insect Molecular Biology, 2002, 11, 329-339), the study of the *ace* genes isolated from other insects or else from arachnids, by homology with that of drosophila, indicates that they are not involved in insecticide resistance.

In fact, no mutation in the amino acid sequence of AChE encoded by the *ace* gene of *Aphis gossypii*, of *Nephotettix cincticeps* and of *Boophilus microplus* is observed between resistant and sensitive individuals (Menozzi et al., doctoral thesis from the Paul Sabatier university, Toulouse, 2000; Tomita et al., Insect Biochem. Mol. Biol., 200, 30, 325-333; Baxter et al., Insect Biochem. Mol. Biol., 1998, 28, 581-589; Hernandez et al., J. Med. Entomol., 1999, 36, 764-770), and independent segregation is observed between the *Culex pipiens* and *C. tritaeniorynchus ace* gene and insecticide resistance (Malcolm et al., Insect. Mol. Biol., 1998, 7, 107-120; Mori et al., Insect Mol. Biol., 2001, 10, 197-203).

As regards the other *ace* genes isolated from other insects, their role in insecticide resistance has not been studied (*Lucilia cuprina*: Chen et al., Insect. Biochem. Mol. Biol., 2001, 31, 805-816; *Schizaphis graminum*: Gao et al., Insect. Biochem. Mol. Biol., 2001, 31, 1095-1104) or no insecticide-insensitive form

of AChE has been described (*Aedes aegypti, Anopheles gambiae* and *Anopheles stephensi*: Anthony et al., FEBS letters, 1995, 368, 461-465; Malcolm et al., in *Molecular Insect Science*, Eds Hageborn et al., Plenum Press, New York, pp 57-65).

Two hypotheses have been put forward to explain the difference in insecticide resistance observed between *Drosophila melanogaster* or *Musca domestica* and the other insects or the arachnids which have been studied: the presence of a "modifier gene" responsible for post-transcriptional or post-translational modifications of AChE, resulting in AChE forms having different catalytic activities, and the presence of a second *ace* gene.

However, no study has made it possible to verify these hypotheses and, consequently, to determine the nature of the gene and that of the target (AChE) involved in insecticide resistance in insects other than *Drosophila melanogaster* and *Musca domestica* or else in arachnids:

- The demonstration, in *C. pipiens*, of two AChE forms having distinct catalytic activities supports the two hypotheses, but the biochemical analysis of these AChEs has not made it possible to determine the nature of the AChE involved in insecticide resistance (Bourguet et al., J. Neurochemistry, 1996, 67, 2115-2123). In fact, the description of an AChE1 activity insensitive to propoxur in insect extracts by Bourguet et al. (Pesticide Biochemistry and Physiology, 1996, 55, 2, 122-128) provides no data on the effective existence of AChE1 in *Culex pipiens*, nor on the separation of AChE1 activity from AChE2 activity, in the context of two hypotheses mentioned above, in light of the subsequent article by the same authors (Bourguet D. et al., Neurochemistry Internat., 1997, 31, 1, 65-72), in which the existence of a second gene in many mosquitoes could not be demonstrated.

- A second *ace* gene has been isolated in arachnids; however, this gene is not involved in

insecticide resistance (Hernandez et al., Baxter et al., mentioned above).

- It has not been possible to isolate a second *ace* gene in insects despite many attempts in various species (Menozzi et al., Tomita et al., Mori et al., mentioned above; Severson et al., J. Hered., 1997, 88, 520-524).

It emerges from the above that the nature of the gene and of the target (AChE) involved in the resistance to organophosphorus compounds and/or to carbamates has not been identified in most insects and in arachnids, in particular in those in which they have been investigated; mention may be made of those which are the most important in the human and animal health fields and agricultural field, such as pathogen vectors and pests, in particular many mosquitoes such as *Culex pipiens*, *Aedes aegypti*, *Anopheles gambiae*, *Anopheles albimanus* or *Anopheles stephensi*, and crop pests such as *Aphis gossypii*, *Nephotettix cincticeps* and *Leptinotarsa decemlineata*.

The inventors have identified a novel locus of the *ace* gene in the genome of *Anopheles gambiae* and of 15 different species of mosquitoes, and they have shown that this novel locus, which is not homologous to the locus previously described in *D. melanogaster*, is involved in insecticide resistance in mosquitoes.

The inventors have also shown that the insecticide resistance, at least in mosquitoes of the species *Culex pipiens* and *Anopheles gambiae*, is linked to a unique mutation in the acetylcholinesterase sequence encoded by this novel gene, located in the region of the catalytic site of the enzyme.

This novel gene represents a diagnostic tool for the genetic detection of insecticide (organophosphorus compound, carbamate) resistance in mosquito populations. The AChE encoded by this gene represents a target for the screening of novel molecules that are active on the populations of mosquitoes resistant to the insecticides currently used.

Consequently, a subject of the present invention is a protein, characterized in that it comprises a central catalytic region which has an amino acid sequence selected from the group consisting of the sequence SEQ ID NO. 1 and the sequences exhibiting at least 60% identity or 70% similarity with the sequence SEQ ID NO. 1, with the exclusion of the NCBI sequence AAK0973 corresponding to the *Schizaphis graminum* acetylcholinesterase.

The protein according to the invention represents a novel insect acetylcholinesterase, hereinafter referred to as AchE1, responsible for resistance to organophosphorus compounds and/or to carbamates, at least in mosquitoes, in particular in *C. pipiens*; the locus encoding said AchE1 is hereinafter referred to as *ace-1*; *ace-2* represents the second *ace* locus, which is not involved in insecticide resistance in mosquitoes. The single *ace* gene present in *Drosophila melanogaster*, which is homologous to *ace-2*, is therefore also referred to as *ace-2*.

In accordance with the invention, said central catalytic region contains the catalytic domain of the AChE, and corresponds to that located between positions 70 and 593 of the sequence of *Anopheles gambiae* AChE1 (SEQ ID NO. 3, 643 amino acids); it corresponds to that located, respectively, between positions 100 and 629 of the sequence of *Schizaphis graminum* AChE1 (NCBI AAK0973), 60 and 582 of the sequence of *Culex pipiens* AChE1 (SEQ ID NO. 7), 34 and 593 of the sequence of *Anopheles gambiae* AChE2 (figure 1, SEQ ID NO. 53), and 41 and 601 of the sequence of *Drosophila melanogaster* AChE2 (NCBI AAF54915). This central region which contains the catalytic domain is conserved in vertebrates and invertebrates, whereas the N- and C-terminal ends exhibit great variability between the various species.

In accordance with the invention, the identity of a sequence relative to a reference sequence (SEQ ID NO. 1) is assessed according to the percentage of amino

acid residues which are identical, when the sequences corresponding to the catalytic region as defined above are aligned, so as to obtain the maximum correspondence between them.

A protein having an amino acid sequence having at least X% identity with the reference sequence SEQ ID NO. 1 is defined, in the present invention, as a protein which sequence corresponding to the central catalytic region as defined above can include up to 100-X alterations per 100 amino acids of the sequence SEQ ID NO. 1. For the purpose of the present invention, the term "alteration" includes consecutive or dispersed amino acid deletions, substitutions or insertions in the reference sequence. This definition applies, by analogy, to the nucleic acid molecules.

The similarity of a sequence relative to the reference sequence SEQ ID NO. 1 is assessed according to the percentage of amino acid residues which are identical or which differ by conservative substitutions, when the sequences corresponding to the central catalytic region as defined above are aligned so as to obtain the maximum correspondence between them. For the purpose of the present invention, the term "conservative substitution" is intended to mean the substitution of an amino acid with another which has similar chemical properties (size, charge or polarity), which generally does not modify the functional properties of the protein.

A protein having an amino acid sequence having at least X% similarity with the sequence SEQ ID NO. 1 is defined, in the present invention as a protein which sequence corresponding to the central catalytic region as defined above can include up to 100-X non-conservative alterations per 100 amino acids of the reference sequence. For the purpose of the present invention, the term "non-conservative alterations" includes consecutive or dispersed amino acid deletions, non-conservative substitutions or insertions in the sequence SEQ ID NO. 1.

The comparison of the AChE1 according to the invention with the insect AChEs available on the databases, by alignment of the sequences corresponding to the central region as defined above, using the BLAST program (http://www.ncbi.nlm.nih.gov/gorf/bl2.html, default parameters, inactivated filter) shows that:

- the insect AChE1 and AChE2 sequences exhibit 36-39% identity (53-57% similarity) with one another,
- the insect AChE1 sequences exhibit 65-97% identity (79-98% similarity) with one another,
- the insect AChE2 sequences exhibit 58-99% identity (73-99% similarity) with one another.

In addition, the phylogenetic analysis of the AChEs of the various animal species shows that the AChE1 protein sequences form a significant autonomous group (bootstrap 795/1000), and that the insect AChE1s form a significant distinct subgroup (bootstrap 856/1000).

The AChE1 according to the invention comprises units characteristic of AChEs (figure 1) located at the following positions, respectively, in the sequence SEQ ID NO. 3 and in the reference sequence from *Torpedo californica* (SWISSPROT P04058): a canonic unit of the FGESAG type around the serine at position 266 (*200*), which is characteristic of the AChE active site, a choline-binding site (tryptophan residue at position 151 (*84*)), three residues of the catalytic triad (serine, glutamic acid and histidine residues, respectively at positions 266 (*200*), 392 (*327*) and 506 (*440*)), six cysteine residues potentially involved in conserved disulfide bridges ($C_{134(67)}$-$C_{161(94)}$; $C_{320(254)}$-$C_{333(265)}$; $C_{468(402)}$-$C_{589(521)}$), aromatic residues bordering the active site gorge (10 residues) and a phenylalanine residue at position 355 (*290*) but not at position 353 (*288*), which distinguishes invertebrate AChEs from those of vertebrates. It also has a hydrophobic C-terminal peptide corresponding to a glycolipid addition signal, indicating post-translational cleavage of a C-terminal fragment and the addition of a glycolipid

anchoring residue as in *Drosophila*; the cysteine residue in the C-terminal sequence preceding the potential site for cleavage of the hydrophobic peptide could be involved in an intermolecular disulfide bond linking the two catalytic subunits of the AChE dimer.

The AChE1 according to the invention differs from the AChE of *Drosophila* (AChE2) by the absence of a hydrophilic insertion of 31 amino acids between the residues located at positions 174 and 175 of the sequence SEQ ID NO. 3 (figure 1); this hydrophilic insertion could be characteristic of AChE2, at least in the Diptera.

The invention encompasses the insect AChE1s sensitive or resistant to organophosphorus compounds and/or to carbamates.

For the purpose of the present invention, the AChE1 sequences include both the primary sequences and the secondary sequences and the tertiary sequences of said AChE1s.

For the purpose of the present invention, the term "sensitive AChE" is intended to mean an AChE for which the acetylcholinesterase activity is inhibited in the presence of organophosphorus compounds or of carbamates.

For the purpose of the present invention, the term "resistant AChE" is intended to mean an AChE for which the activity is not inhibited by concentrations of organophosphorus compounds or of carbamates which inhibit 100% of the activity of the corresponding "sensitive AChE" derived from an individual of the same species; this "resistant AChE" differs from the preceding one by the presence of one or more mutations in its amino acid sequence (amino acid substitutions) which modify its sensitivity to acetylcholinesterase inhibitors; among these mutations, mention may be made of the following: F78S, I129V, G227A and F228Y, the amino acids being numbered with reference to the sequence of *Torpedo californica* AChE (SWISSPROT P04058).

The acetylcholinesterase activity and the catalytic parameters of the AChEs are measured by conventional enzymatic techniques such as those described in Bourguet et al., mentioned above.

The proteins according to the invention include any natural, synthetic, semi-synthetic or recombinant protein of any prokaryotic or eukaryotic organism, comprising or consisting of an amino acid sequence of an AChE1 protein as defined above. They include in particular the natural proteins isolated from any insect species, and also the recombinant proteins produced in a suitable expression system.

According to an advantageous embodiment of said AChE1, it corresponds to that of an insect which belongs to the order *Diptera*; preferably, said insect is chosen from the family *Culicidae*, from the genera *Culex*, *Aedes* and *Anopheles*.

According to an advantageous arrangement of this embodiment, said AChE1 consists of the sequences SEQ ID NO. 3, SEQ ID NO. 5 and SEQ ID NO. 126 of *Anopheles gambiae* and of sequence SEQ ID NO. 7 of *Culex pipiens* (S-LAB strain), that are sensitive to organophosphorus compounds and/or to carbamates.

According to another advantageous arrangement of this embodiment, said central catalytic region of the AChE1 corresponds to a sequence selected from the group consisting of the sequences SEQ ID NOs. 8 to 21 representing a fragment of approximately 91 amino acids (fragment K, figure 1), corresponding to that located between positions 445 and 535 of the sequence SEQ ID NO. 3.

According to another advantageous embodiment of the invention, said AChE1 is an acetylcholinesterase resistant to insecticides of the organophosphorus compound and carbamate class, that includes a mutation of the glycine located at position 119 to serine (mutation or substitution of G119S type), said position being indicated with reference to the sequence of *Torpedo californica* AChE (SWISSPROT P04058).

In fact, the inventors have shown that the residue at position 119 is close to the residues of the catalytic site (serine 200 and histidine 440) and that the replacement of the glycine of the AChE1 of sensitive mosquitoes with a serine, in the AChE1 of resistant mosquitoes, reduces the space of the catalytic site and prevents the insecticide from interacting with the catalytic serine (S200), due to the steric hindrance of the Van der Waals bonds of the side chain of the serine at position 119. The role of the G119S mutation in the insecticide resistance has been confirmed by analysis of the acetylcholinesterase activity of recombinant AChE1 proteins produced from the cDNA of *Culex pipiens* sensitive (S-LAB strain having an AChE1 that includes a glycine at position 119) or resistant (SR strain in which the AChE1 differs from the preceding one only by the presence of a serine at position 119) to insecticides; 90% of the activity of the AChE1 of the sensitive strain is inhibited in the presence of 10^{-3} M of propoxur, whereas the AChE1 of the resistant strain conserves 75% of its activity in the presence of 100-times higher concentrations of this insecticide (10^{-1} M of propoxur).

According to an advantageous arrangement of this embodiment of said resistant AChE1, it corresponds to that of an insect (resistant to insecticides) which belongs to the order *Diptera*; preferably, said insect is chosen from the family *Culicidae*, from the genera *Culex*, *Aedes* and *Anopheles*.

Preferably, said resistant AChE1 has a sequence selected from the group consisting of:

- the sequence SEQ ID NO. 57, corresponding to the complete sequence of the SR strain of *C. pipiens*, resistant to insecticides,

- the sequence SEQ ID NO. 122, corresponding to the complete sequence of the AChE1 of the YAO strain of *An. gambiae* (isolated in Ivory Coast), resistant to insecticides, and

- the sequences comprising a fragment of sequence SEQ ID NOs. 90, 93, 94, 95, 97 to 101, 113 and 116 representing a peptide fragment of approximately 150 amino acids encoded by the third coding exon of the *ace-1* gene of a resistant insect as defined above, containing the substitution of G119S type, with reference to the sequence of *Torpedo californica* AChE (SWISSPROT P04058).

According to yet another advantageous embodiment of the invention, said AChE1 is an acetylcholinesterase sensitive to insecticides of the organophosphorus compound and carbamate class, comprising a sequence selected from the group consisting of SEQ ID NOs. 91, 92, 96, 102 to 112, 114, 115 and 117 to 119, representing a fragment of approximately 150 amino acids of the third coding exon of the *ace-1* gene derived from an insect as defined above, sensitive to insecticides, said fragment including a glycine at position 119 with reference to the sequence of *Torpedo californica* AChE (SWISSPROT P04058).

A subject of the present invention is also a peptide, characterized in that it consists of a fragment of at least 7 amino acids of the AChE1 protein, as defined above; these fragments are particularly useful for producing antibodies that specifically recognize the AChE1 protein.

A subject of the present invention is also antibodies, characterized in that they are directed against the AChE1 protein or a fragment thereof, as defined above.

In accordance with the invention, said antibodies are either monoclonal antibodies or polyclonal antibodies.

These antibodies may be obtained by conventional methods, known in themselves, comprising in particular the immunization of an animal with a protein or a peptide in accordance with the invention, in order to make it produce antibodies directed against said protein or said peptide.

A subject of the present invention is also an isolated nucleic acid molecule, characterized in that it has a sequence selected from the group consisting of:

- the sequences encoding an AChE1 protein as defined above (cDNA and genomic DNA fragment corresponding to the *ace-1* gene),
- the sequences complementary to the above sequences, which may be sense or antisense, and
- the fragments of at least 8 bp, preferably of 15 bp to 500 bp, of the above sequences.

The invention encompasses the sequences of the alleles of the *ace-1* gene derived from any insect, and also the sequences of the natural (sensitive or resistant alleles) or artificial mutants of the *ace-1* gene encoding a sensitive or resistant AChE1 protein, as defined above.

According to an advantageous embodiment of the invention, said sequence encoding an AChE1 protein is selected from the group consisting of:

a) the sequences SEQ ID NO. 2, SEQ ID NO. 4, SEQ ID NO. 125, SEQ ID NO. 6, SEQ ID NO. 56 and SEQ ID NO. 121 which correspond to the cDNA of the AChE1 protein of amino acid sequence, respectively, SEQ ID NO. 3, SEQ ID NO. 5, SEQ ID NO. 126, SEQ ID NO. 7, SEQ ID NO. 57 and SEQ ID NO. 122, as defined above,

b) the sequences SEQ ID NO. 22, SEQ ID NO. 23 and SEQ ID NO. 127 which correspond to the *ace-1* gene of *Anopheles gambiae* encoding the AChE1s as defined above, which gene has an exon-intron organization comprising at least 9 exons (table I), and

c) the sequences comprising the sequence SEQ ID NO. 120 which corresponds to the virtually complete sequence of the *ace-1* gene of *Anopheles gambiae* encoding the resistant AChE1 of sequence SEQ ID NO. 122, as defined above.

Table I: Intron-exon organization of the *ace-1* gene

	5' site		3' site	
	Position	Sequence	Position	Sequence
Intron1	301	AGCAA/gtaat	1255	cgcag/CCATT
Intron2	1413	CAATG/gtgag	5338	tgtag/CGCTC
Intron3	5696	CGCAG/gtcgg	7634	ttcag/ACGCA
Intron4	7769	CTCGG/gtaag	7855	ggcag/ACGCG
Intron5	8393	CTACG/gtagg	8472	gtcag/CTGGG
Intron6	8670	CTAAG/gtacg	8756	tccag/AGCAC
Intron7	9464	ACCGG/gtaag	9530	tacag/CAATC
Intron8	9703	TACCT/gtaag	9810	aacag/CGAAC

In accordance with the present invention, the third coding exon of the *ace-1* gene corresponds to that which is located between intron 4 and intron 5 in the sequence of *An. gambiae* (table I), i.e. between positions 7854 and 8393 of the sequence SEQ ID NO. 127.

According to another advantageous embodiment of the invention, said fragment is selected from the group consisting of the primers of sequence SEQ ID Nos. 39 to 50, 54, 55, 58, 59, 123, 124, 128 and 129 and the fragments of sequences SEQ ID Nos. 24 to 38 and 60 to 89.

The nucleic acid molecules according to the invention are obtained by conventional methods, known in themselves, according to standard protocols such as those described in *Current Protocols in Molecular Biology (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA)*. For example, they can be obtained by amplification of a nucleic acid sequence by PCR or RT-PCR, by screening genomic DNA libraries by hybridization with a homologous probe, or else by total or partial chemical synthesis.

The nucleic acid molecules as defined above can be used as probes or as primers for isolating the *ace-1* gene of other species or alleles of this gene, in particular by screening a genomic DNA or cDNA library, and also for detecting/amplifying nucleic acid

molecules (mRNA or genomic DNA) encoding an AChE1 protein as defined above.

These various nucleic acid molecules make it possible to demonstrate the *ace-1* gene, allelic variants of this gene, or a functional alteration of this *ace-1* gene (substantial change in insecticide sensitivity) resulting from a mutation (insertion, deletion or substitution) of one or more nucleotides in said gene.

A subject of the present invention is also a method for detecting insects carrying resistance to insecticides of the organophosphorus compound and carbamate class, characterized in that it comprises:

- preparing a sample of nucleic acids from insects to be tested, and

- detecting, by any suitable means, the presence, in said nucleic acid sample, of a mutation in the *ace-1* gene as defined above.

Said detection is carried out by conventional techniques which are known in themselves, for example: (i) by amplification of a region of said *ace-1* gene liable to contain a mutation, and then detection of said mutation by sequencing, or by digestion with a suitable restriction enzyme, of the PCR product obtained, or else (ii) by hybridization with a labeled probe specific for a region of said *ace-1* gene liable to contain a mutation, and then direct detection of the mismatches and/or digestion with a suitable restriction enzyme.

According to a first advantageous embodiment of said method, a fragment of approximately 320 bp (fragment K) is amplified by means of the primers SEQ ID NO. 39 and SEQ ID NO. 40. For example, in mosquitoes, a fragment of sequence SEQ ID NOs. 24 to 38 is obtained, which has mutations between insecticide-sensitive and insecticide-resistant mosquitoes. For example, in *C. pipiens* 3 substitutions are observed in the sequence of the resistant individuals, one of which introduces an *EcoRI* site. Analysis of the restriction

profile after PCR amplification of the fragment K and digestion of the products obtained with *EcoRI* (RFLP analysis) makes it possible to rapidly detect the *ace-1* genotype in a population of *C. pipiens*; the presence of a single fragment corresponds to the resistant homozygotes (RR), the presence of 2 fragments of approximately 106 bp and 214 bp corresponds to the sensitive homozygous individuals (SS) and the presence of 3 fragments of 106 bp, 214 bp and 320 bp corresponds to the resistant heterozygous individuals (RS).

According to a second advantageous embodiment of said method, the G119S mutation in the third coding exon of the *ace-1* gene which is responsible for the resistance to insecticides of the organophosphorus compound and carbamate class in mosquitoes is detected according to one of the following alternatives, respectively in mosquitoes of the species *C. pipiens* and *An. gambiae*:

- in mosquitoes of the species *Culex pipiens*, a 520 bp fragment of the third coding exon is amplified from the genomic DNA by PCR using the pair of primers Ex3dir and Ex3rev (SEQ ID NOs. 58 and 59); the PCR fragment is digested with *Alu I* and the digestion product is separated by agarose gel electrophoresis, and then the restriction profile thus obtained is analyzed: the presence of a 520 bp fragment corresponds to the sensitive homozygous SS individuals, the presence of two fragments (357 bp and 163 bp) corresponds to the resistant homozygous RR individuals and the presence of 3 fragments (520 bp, 357 bp and 163 bp) corresponds to the resistant heterozygous RS individuals,

- in mosquitoes of the species *Anopheles gambiae*, a 541 bp fragment of the third coding exon is amplified from the genomic DNA by PCR using the pair of primers Ex3AGdir and Ex3AGrev (SEQ ID NOs. 123 and 124); the PCR fragment is digested with *Alu I* and the digestion product is separated by agarose gel electrophoresis, and then the restriction profile thus obtained

is analyzed: the presence of two fragments (403 bp and 138 bp) corresponds to the sensitive homozygous SS individuals, the presence of 3 fragments (253 bp, 150 bp and 138 bp) corresponds to the resistant homozygous RR individuals and the presence of 4 fragments (403 bp, 253 bp, 150 bp and 138 bp) corresponds to the resistant heterozygous RS individuals; given that the 150 bp and 138 bp fragments co-migrate, the resistant homozygous and heterozygous individuals are detected, respectively, by the presence of 2 bands (253 bp and approximately 150 bp) and of 3 bands (403 bp, 253 bp and approximately 150 bp),

- in mosquitoes of the species *Culex pipiens*, *Anopheles gambiae* or *Anopheles albimanus*, a 194 bp fragment containing codon 119 of the third coding exon is amplified from the genomic DNA by PCR using the pair of primers Moustdir1 and Moustrev1 (SEQ ID NOs. 128 and 129); the PCR fragment is digested with *Alu I* and the digestion product is separated by agarose gel electrophoresis, and then the restriction profile thus obtained is analyzed: the presence of two fragments (74 bp and 120 bp) corresponds to the resistant homozygous RR individuals, the presence of a single fragment (no digestion) corresponds to the sensitive homozygous SS individuals and the presence of three fragments (194 bp, 74 bp and 120 bp) corresponds to the resistant heterozygous RS individuals.

A subject of the present invention is also a reagent for detecting insects carrying resistance to organophosphorus compounds and/or to carbamates, characterized in that it is selected from the group consisting of: the nucleic acid molecules and the fragments thereof as defined above (probes, primers) and the antibodies as defined above.

A subject of the present invention is also a recombinant vector, characterized in that it comprises an insert selected from the group consisting of the nucleic acid molecules encoding an AChE1 protein and the fragments thereof as defined above.

Preferably, said recombinant vector is an expression vector in which said nucleic acid molecule or one of its fragments are placed under the control of suitable regulatory elements for transcription and for translation.

These vectors are constructed and introduced into host cells by conventional recombinant DNA and genetic engineering methods which are known in themselves. Many vectors into which a nucleic acid molecule of interest may be inserted in order to introduce it into and to maintain it in a eukaryotic or prokaryotic host cell are known in themselves; the choice of a suitable vector depends on the use envisioned for this vector (for example, replication of the sequence of interest, expression of this sequence, maintenance of the sequence in extrachromosomal form or else integration into the host's chromosomal material), and also on the nature of the host cell. For example, viral vectors such as baculoviruses or nonviral vectors such as plasmids may be used. In order to express the AChE1, the *ace-1* cDNA may be placed under the control of a constitutive promoter such as the actin 5C promoter, in a suitable vector, and said recombinant vector is introduced into insect cells such as drosophila cells (Schneider S2 cells).

A subject of the present invention is also prokaryotic or eukaryotic cells modified with a recombinant vector as defined above; preferably, these cells are insect cells.

The recombinant vectors and the modified cells as defined above are useful in particular for producing the AChE1 proteins and peptides according to the invention.

A subject of the present invention is also a transgenic invertebrate animal, characterized in that it contains cells modified with at least one nucleic acid molecule as defined above; preferably, said animal is an insect.

The transgenic animals and the modified cells as defined above are useful in particular for screening insecticidal substances and for biologically controlling pathogen vectors and insect pests.

A subject of the present invention is also a method for screening an insecticidal substance, characterized in that it comprises:

a) bringing the test substance into contact with an AChE1 protein selected from: an AChE1 protein isolated according to the invention, or an extract of modified cells or a biological sample from a transgenic animal containing said AChE1 protein, as defined above, in the presence of acetylcholine or of one of its derivatives,

b) measuring, by any suitable means, the acetylcholinesterase activity of the mixture obtained in a), and

c) selecting the substances capable of inhibiting said activity.

A subject of the present invention is also a method for screening an insecticidal substance, characterized in that it comprises:

- bringing a transgenic animal as defined above into contact with the test substance, and
- measuring the animal's survival.

Advantageously, said screening methods use AChE1s resistant to organophosphorus compounds or to carbamates, or else cells or transgenic animals containing them.

A subject of the present invention is also a reagent for screening insecticidal substances, characterized in that it is selected from the group consisting of the AChE1 proteins, the recombinant vectors, the modified cells and the transgenic animals as defined above.

Insecticidal substances capable of inhibiting the acetylcholinesterase activity of the AChE1 proteins resistant to insecticides of the organophosphorus compound and carbamate class commonly used have appli-

cations: in animal and human health, for controlling pathogen vectors (for example *Aedes aegypti*, a vector of arboviroses such as dengue and yellow fever, *Culex pipiens*, a West-Nile virus vector, *Anopheles gambiae*, an African vector of the agent for malaria, etc.) and in the agricultural field, for controlling insect pests which devastate harvests (for example the Colorado potato beetle (*Leptinotarsa decemlineata*) which attacks potatoes, aphid pests such as *Aphis gossypii* and *Myzus persicae*, etc.).

A subject of the invention is also a detection and/or screening kit for carrying out the methods as defined above, characterized in that it includes at least one reagent as defined above.

A subject of the present invention is also a method for screening inhibitors of an AChE1 as defined above, characterized in that it comprises:

(a) identifying molecules (peptides or other chemical structures) having a significant probability of binding to said AChE1;

(b) isolating the potential inhibitors identified in step (a);

(c) bringing the substance isolated in step (b) into contact with an AChE1 as defined above, an extract of modified cells, a biological sample from a transgenic animal as defined above, or an extract of an insect sensitive or resistant to the abovementioned insecticides, in the presence of acetylcholine or of one of its derivatives;

(d) measuring, by any suitable means, the acetylcholinesterase activity of the mixture obtained in (c); and

(e) verifying that the molecules isolated in (b) inhibit the AChE1 activity.

The 3D structure of the torpedo fish acetylcholinesterase has made it possible to model the 3D structure of the *C. pipiens* AChE1. The G247S mutation [corresponding to a G119S substitution in the torpedo protein] results in a decrease in the space of the

catalytic site due to the hindrance of the serine side chain.

Modeling of the structure of the *C. pipiens* or *An. gambiae* AChE1 thus makes it possible to screen AChE1 inhibitors by virtual screening ("Docking"). The method according to the invention comprises a computer simulation step (step (a)) aimed at identifying peptide or chemical structures having a significant probability of binding to a target protein. Various computer programs thus make it possible to simulate and to estimate the probabilities of interactions. Mention may in particular be made of the algorithms developed for the search for potential interactions described in Schneider et al. (Drug Discovery Today, 2002, 7, 1, 64-71). More precisely, the tools most commonly used up until now are FlexX (Tripos, St Louis, Missouri, USA), DOCK (UCSF, San Francisco, California, USA) and GOLD (Cambridge Crystallographic Data Centre, Cambridge, United Kingdom).

It is thus possible to isolate potential inhibitors of the resistant form of AChE1 without biochemically having the protein, and then to directly test the inhibitory ability of each candidate on the AChE1 activity of an extract of sensitive or resistant insects (step (c) of the method). This approach can therefore be entirely free of the purification and/or of the production of target protein.

For the purpose of the present invention, the significance of a binding probability cannot be defined absolutely: it may depend on the type of amino acids involved in the interaction, and also on the programs used for the modeling. More precisely, the most commonly used methods select, for a given site of a target molecule, the compounds exhibiting the lowest binding energy. In general, the calculation of the energy takes into account the "hydrogen" bonds, the van der Waals, electrostatic and hydrophobic interactions, and also the entropy penalties. It is therefore *a priori* impossible to give a significant limit in

absolute value beyond which a potential ligand will be accepted or rejected, since the energy will depend on the atoms involved in the bond. However, three selection criteria can be applied:

1. an arbitrary selection of the compounds with least binding energy. In general, the limit is fixed between 1% and 5% of the number of compounds tested.

2. an estimation of the binding affinity, according to the energy calculations. An acceptable value as a starting point could be between 1 and 300 micromolar. By way of example, onchidal, an AChE inhibitor, has an apparent affinity of 300 µM (Abramson et al., *Mol. Pharmacol.*, 1989, **36**, 349).

3. a statistical selection of the compounds, estimating the probability of an identical or higher score for a compound being obtained randomly. In general, the compound is accepted when the estimated probability is ≤ 0.05.

Besides the above provisions, the invention also comprises other provisions which will emerge from the following description, which refers to examples of implementation of the *ace-1* gene and of its products (cDNA, protein) according to the present invention and also to the table summarizing the sequences of the application and to the attached drawings in which:

- figure 1 illustrates the amino acid sequence alignment for the AChE1 proteins of *Anopheles gambiae, Schizaphis graminum, An. stephensi, Aedes aegypti, Drosophila melanogaster, Lucilia cuprina, Musca domestica* and *Culex pipiens*. By convention, the amino acids are numbered with reference to the AChE sequence from the torpedo fish (*Torpedo californica*; SWISSPROT P04058). The N- and C-terminal sequences are not represented due to their variability. The amino acids that are conserved between AChE1 and AChE2 are indicated in gray. The amino acids specific for AChE2 are indicated in black. The 3 residues representing the catalytic triad (S_{200}, E_{327} and H_{440}) have boxes drawn round them. The choline-binding site (W_{84}) is under-

lined. The circles represent the position of the 14 aromatic residues bordering the active site gorge in the *Torpedo* AChE, 10 of which are present in all the AChE1s and AChE2s (solid circles), the others not being conserved (open circles). Three intramolecular disulfide bonds between cysteine residues are indicated. The horizontal arrow indicates the position of the fragment K (amplified by means of the primers PdirAGSG and PrevAGSG). The hypervariable region of AChE2, which is absent in AChE1, is enclosed in a box;

- figure 2 illustrates the genetic detection of mosquitoes resistant to organophosphorus compounds and/or to carbamates by PCR-RFLP:

- figure 2A represents the comparison of the amino acid sequence of fragment K of various mosquito species: Cx Pip (*Culex pipiens*), Ae alb (*Aedes albopicus*), Ae aeg (*Aedes aegypti*), An alb (*Anopheles albimanus*), An gamb (*Anopheles gambiae*), An fun (*Anopheles funestus*), An nil (*Anopheles nili*), An sac (*Anopheles sacharovi*), An pse (*Anopheles pseudopunctipennis*). The variant amino acids are shaded. The following sequences are identical: *An. darlingi* and *An. albimanus*; *An. sundaicus*, *An. gambiae* and *An. arbiensis*; *An. moucheti*, *An. funestus* and *An. minimus*; *An. stephensi* and *An. saccharovi*;

- figure 2B illustrates the comparison of the nucleotide sequences corresponding to fragment K of the sensitive (S-LAB) and resistant (SR) strains. The variant nucleotides are shaded (t → c at position 3; a → g at position 84: the *EcoRI* site (gaattc) located around this position, used for the PCR-RFLP analysis, is present only in the S-LAB strain; c → t at position 173). Figure 2C illustrates the restriction profiles obtained after agarose gel electrophoresis of the products of digestion with *EcoRI*, of fragment K amplified by PCR. The sensitive homozygous S-LAB strain has a profile characterized by 2 bands (214 bp and 106 bp), the resistant homozygous strain has a profile characterized by a single band of 320 bp and the

resistant mosquitoes derived from the back cross have a heterozygous profile characterized by 3 bands (320 bp, 214 bp and 106 bp);

- figure 3 illustrates the phylogenetic tree for the AChE proteins. The phylogenetic analysis was carried out using 47 sequences of AChE proteins from 35 different species originating from the ESTHER database (http://www.ensam.inra.fr/cgi-bin/ace/index). The sequences were aligned and a tree was constructed as described in example 1. Only the nodes corresponding to "bootstrap" values > 50% (i.e. scores greater than 500) are indicated. The scale represents a divergence of 10%. Agam: *An. gambiae*; Aeg: *Aedes aegypti*; Aste: *Anopheles stephensi*; Cp: *Culex pipiens*; Dme1: *Drosophila melanogaster*; Lcup: *Lucilia cuprina*; Mdom: *Musca domestica*; Ldec: *Leptinotarsa decemlineata*; Ame1: *Apis mellifera*; Ncin: *Naphotettix cincticeps*; Sgra: *Schizaphis graminum*; Rapp: *Rhipicephalus appendiculatus*; Bmic: *Boophilus microplus*; Bdec: *Boophilus decoloratus*; Hsap: *Homo sapiens*; Btau: *Bos taurus*; Fcat: *Felix catus*; Ocun: *Oryctolagus cuniculus*; Rnor: *Rattus norvegicus*; Mmus: *Mus musculus*; Ggal: *Gallus gallus*; Drer: *Danio reno*; Eele: *Electrophorus electricus*; Tamr: *Torpedo marmorata*; Tcal: *Torpedo californica*; Bfas: *Bungarus fasciatus*; Mglu: *Myxine glutinosa*; Bflo: *Branchiostoma floridae*; Blan: *Branchiostoma lanceolatum*; Cint: *Ciona intestinalis*; Csav: *Ciona savignyi*; Cele: *Caenorhabditis elegans*; Cbrig: *Caenorhabditis briggsae*; Dviv: *Dictyocaulus viviparus*; Lopa: *Loligo opalescens*;

- figure 4 illustrates the cladogram for the AChE1 and AChE2 proteins. The sequences of the AChE1 and AChE2 proteins were processed as in figure 1. The Bmic sequence was added as external sequence in order to define the origin of the tree. The boxes marked with an asterisk represent the proteins encoded by a gene which segregates with insecticide resistance. The open boxes represent the proteins encoded by a gene which does not segregate with insecticide resistance. The scale corresponds to a divergence of 10%;

- figure 5 illustrates the comparison of the amino acid sequences of the AChE1 protein of *C. pipiens*, between an insecticide-sensitive strain (S-LAB) and an insecticide-resistant strain (SR). The single mutation glycine $_{247(119)}$ → serine $_{247(119)}$ (indicated with shading) is responsible for the insecticide resistance in mosquitoes of the species *C. pipiens*; it corresponds to the substitution of the glycine located at position 247 of the sequence of *C. pipiens* AChE1 (or at position 119, with reference to the sequence of the torpedo fish AChE), with a serine;

- figures 6A and 6B illustrate the comparison of the nucleotide sequences encoding the *C. pipiens* AChE1 protein, between an insecticide-sensitive strain (S-LAB) and an insecticide-resistant strain (SR); all the mutations are silent, with the exception of the mutation at position 739 (G → A), which results, firstly, in the substitution of the glycine codon (GGC) at position 247 of the sequence of the AChE1 protein of the sensitive strain (S-LAB) with a serine codon (AGC) responsible for insecticide resistance in the SR strain and, secondly, in the appearance of an *Alu I* site (AGCT) in the sequence of the resistant strain, that is useful for detecting the mutation. The mutation (G → A) at position 739 of the nucleotide sequence and the mutation glycine → serine at position 247 of the amino acid sequence are indicated with shading. The sequences of the primers used for detecting the mutation at position 739 (primer Ex3dir and Ex3rev), and also the *Alu I* site are indicated in bold and are underlined;

- figure 7 (A, B and C) illustrates the three-dimensional structure of the *C. pipiens* AChE1, obtained by molecular modeling from the structure of the torpedo fish AChE:

figure 7A illustrates (i) the overall structure of the two proteins and (ii) the steric hindrance of the Van der Waals bonds of the serine 200 and of the histidine 440 of the catalytic site of the enzyme, and

also that of the amino acid in position 119 which is mutated in cases of resistance; the residue at position 119 is close to the residues S_{200} and H_{440} of the catalytic site;

figures 7B and 7C illustrate the comparison of the steric hindrance of the Van der Waals bonds of the amino acid glycine (figure 7C) and serine (figure 7B) at position 119, of respectively the sensitive and resistant strain. The hindrance of the side chain of the serine at position 119 in the resistant strain decreases the space of the catalytic site, which probably prevents the insecticide from interacting with the catalytic serine (S_{200});

- figure 8 illustrates the detection by PCR-RFLP of the mutation glycine → serine in the third coding exon of the *ace-1* gene, in mosquitoes of the species *C. pipiens*: 1 band (520 bp) is detected in the sensitive homozygous SS individuals, 2 bands (357 bp and 163 bp) are detected in the resistant homozygous RR individuals and 3 bands (520 bp, 357 bp and 163 bp) are detected in the resistant heterozygous RS individuals;

- figures 9A and 9B illustrate the comparison of the sequences of the *An. gambiae ace-1* gene, between an insecticide-sensitive strain (KISUMU) and an insecticide-resistant strain (YAO); all the mutations are silent, with the exception of two mutations: the first corresponds to the replacement of the valine (CGT) at position 33 of the sequence of the AChE1 of the sensitive strain (SEQ ID NO. 5) with an alanine (CGC) in the resistant strain, and the second is the same mutation, glycine (GGC) → serine (AGC) as that found in *Culex pipiens*. The mutation glycine (GGC) → serine (AGC) results in the appearance of a second *Alu I* site (AGCT) in the sequence of the third coding exon of the resistant strain, that is useful for detecting the mutation. The coding sequences of the *ace-1* gene are indicated in bold and the mutations are indicated with shading. The sequences of the primers Ex3AGdir and Ex3Agrev used for detecting the mutation

glycine (GGC) → serine (AGC), and also the *Alu I* sites of the third coding exon, are indicated in bold and are underlined;

- figure 10 illustrates the quantification of the acetylcholinesterase activity of the recombinant *Culex pipiens* AChE1 proteins that are sensitive (S-LAB, white bars) and resistant (SR, shaded bars), produced in S2 insect cells, by comparison with that of ground material from *C. pipiens* strain S-LAB (hatched white bars) and strain SR (hatched shaded bars). The acetylcholinesterase activity of the cell extracts and of the ground materials from mosquitoes was measured in the absence (C) and in the presence of 10^{-4} M and 10^{-2} M of propoxur. The single mutation $glycine_{247(119)}$ → $serine_{247(119)}$ renders the acetylcholinesterase insensitive to the insecticide;

- figure 11 illustrates the detection by PCR-RFLP of the glycine → serine mutation in the third coding exon of the *ace-1* gene, in mosquitoes of the species *Culex pipiens*, *Anopheles gambiae* and *Anopheles albimanus*: 1 band (194 bp) is detected in the sensitive homozygous SS individuals, 2 bands (74 bp and 120 bp) are detected in the resistant homozygous RR individuals and 3 bands (194 bp, 74 bp and 120 bp) are detected in the resistant heterozygous SS individuals;

- figure 12 illustrates the quantification of the acetylcholinesterase activity of the AChE1 proteins of, respectively, *Culex pipiens*, *Anopheles gambiae* and *Anopheles albimanus* that are sensitive (SS, shaded bars) and resistant (RS, black bands and RR, white bands);

- figure 13 represents the alignment of the nucleotide sequences of the 194 bp fragment from *Anopheles gambiae*, from *Culex pipiens* and from *Anopheles albimanus*, that are sensitive (S) or resistant (R). Light shaded background: sequences corresponding to the primers Moustdir1 and Moustrev1. Shaded background: *Alu I* site. Dark shaded background: guanine of the Gly codon of sensitive individuals;

- figure 14 represents the nucleotide sequences of the 194 bp fragment of sensitive (S) and resistant (R) *Anopheles albimanus*. The codon specifying Gly (GGC) and Ser (AGC) is in bold. The *Alu I* site is underlined.

Table II: Sequence listing

Identification number	Sequence
SEQ ID No.: 1	Fragment of the central region of the *Anopheles gambiae* AChE1 protein (positions 70 to 593 of SEQ ID No. 3).
SEQ ID No.: 2	*Anopheles gambiae* AChE1 cDNA
SEQ ID No.: 3	*Anopheles gambiae* AChE1 protein
SEQ ID No.: 4	*Anopheles gambiae* (strain KISUMU) AChE1 cDNA
SEQ ID No.: 5	*Anopheles gambiae* (strain KISUMU) AChE1 protein
SEQ ID No.: 6	*Culux pipiens* strain S-LAB AChE1 cDNA (complete sequence)
SEQ ID No.: 7	*Culex pipiens* strain S-LAB AChE1 protein (complete sequence)
SEQ ID No.: 8	Peptide fragment K AChE1 *Culex pipiens*
SEQ ID No.: 9	Peptide fragment K AChE1 *Aedes aegypti*
SEQ ID No.: 10	Peptide fragment K AChE1 *Aedes albopictus*
SEQ ID No.: 11	Peptide fragment K peptide AChE1 *Anopheles darlingi*
SEQ ID No.: 12	Peptide fragment K AChE1 *An. sundaicus*
SEQ ID No.: 13	Peptide fragment K AChE1 *An. minimus*
SEQ ID No.: 14	Peptide fragment K AChE1 *An. moucheti*
SEQ ID No.: 15	Peptide fragment K AChE1 *An. arabiensis*
SEQ ID No.: 16	Peptide fragment K AChE1 *An. funestus*
SEQ ID No.: 17	Peptide fragment K AChE1 *An. pseudopunctipennis*
SEQ ID No.: 18	Peptide fragment K AChE1 *An. sacharovi*
SEQ ID No.: 19	Peptide fragment K AChE1 *An. stephensi*
SEQ ID No.: 20	Peptide fragment K AChE1 *An. albimanus*
SEQ ID No.: 21	Peptide fragment K AChE1 *An. nili*
SEQ ID No.: 22	*An. gambiae ace-1* gene
SEQ ID No.: 23	*An. gambiae* KISUMU *ace-1* gene
SEQ ID No.: 24	Nucleotide fragment K AChE1 *C. pipiens* (strain S-LAB)
SEQ ID No.: 25	Nucleotide fragment K AChE1 *C. pipiens* (strain SR)
SEQ ID No.: 26	Nucleotide fragment K AChE1 *Aedes aegypti*
SEQ ID No.: 27	Nucleotide fragment K AChE1 *Aedes albopictus* (AJ 438598)
SEQ ID No.: 28	Nucleotide fragment K AChE1 *Anopheles darlingi* (AJ 438599)
SEQ ID No.: 29	Nucleotide fragment K AChE1 *An. sundaicus* (AJ 438600)
SEQ ID No.: 30	Nucleotide fragment K AChE1 *An. minimus* (AJ 438601)
SEQ ID No.: 31	Nucleotide fragment K AChE1 *An. moucheti* (AJ 438602)
SEQ ID No.: 32	Nucleotide fragment K AChE1 *An. arabiensis* (AJ 438603)
SEQ ID No.: 33	Nucleotide fragment K AChE1 *An. funestus* (AJ 438604)
SEQ ID No.: 34	Nucleotide fragment K AChE1 *An. pseudopunctipennis* (AJ 538605)
SEQ ID No.: 35	Nucleotide fragment K AChE1 *An. sacharovi* (AJ 538606)

SEQ ID No.: 36	Nucleotide fragment K AChE1 *An. stephensi* (AJ 538607)
SEQ ID No.: 37	Nucleotide fragment K AChE1 *An. albimanus* (AJ 538608)
SEQ ID No.: 38	Nucleotide fragment K AChE1 *An. nili* (AJ 538609)
SEQ ID No.: 39	Primer PkdirAGSG
SEQ ID No.: 40	Primer PkrevAGSG
SEQ ID No.: 41	Primer PbdirAGSG
SEQ ID No.: 42	Primer PbrevAGSG
SEQ ID No.: 43	Primer culex-bdir1
SEQ ID No.: 44	Primer culex-krev1
SEQ ID No.: 45	Primer AG1-Adir
SEQ ID No.: 46	Primer AG1-Arev
SEQ ID No.: 47	Primer AG1-Bdir
SEQ ID No.: 48	Primer AG1-Brev
SEQ ID No.: 49	Primer AG1-Cdir
SEQ ID No.: 50	Primer AG1-Crev
SEQ ID No.: 51	*Ciona intestinalis* AChE1 protein
SEQ ID No.: 52	*Ciona savignyi* AChE1 protein
SEQ ID No.: 53	*Anopheles gambiae* AChE2 protein
SEQ ID No.: 54	Primer culex-5'dir
SEQ ID No.: 55	Primer culex-3'dir
SEQ ID No.: 56	*C. pipiens* strain SR AChE1 cDNA (complete coding sequence)
SEQ ID No.: 57	*C. pipiens* strain SR AChE1 protein (complete sequence)
SEQ ID No.: 58	Primer Ex3dir
SEQ ID No.: 59	Primer Ex3rev
SEQ ID No.: 60	Nucleotide fragment of the third coding exon strain Espro-P*-R****
SEQ ID No.: 61	Nucleotide fragment of the third coding exon strain Pro-R-Q**-S
SEQ ID No.: 62	Nucleotide fragment of the third coding exon strain S-LAB-Q-S*****
SEQ ID No.: 63	Nucleotide fragment of the third coding exon strain Padova-P-R
SEQ ID No.: 64	Nucleotide fragment of the third coding exon strain Praias-P-R
SEQ ID No.: 65	Nucleotide fragment of the third coding exon strain Supercar-Q-R
SEQ ID No.: 66	Nucleotide fragment of the third coding exon strain BrugesA-P-S
SEQ ID No.: 67	Nucleotide fragment of the third coding exon strain BQ-Q-R
SEQ ID No.: 68	Nucleotide fragment of the third coding exon strain DJI-Q-R
SEQ ID No.: 69	Nucleotide fragment of the third coding exon strain Harare-Q-R
SEQ ID No.: 70	Nucleotide fragment of the third coding exon strain Martinique-Q-R
SEQ ID No.: 71	Nucleotide fragment of the third coding exon strain Barriol-P-R
SEQ ID No.: 72	Nucleotide fragment of the third coding exon strain Bleuet-P-S
SEQ ID No.: 73	Nucleotide fragment of the third coding exon strain BrugesB-P-S
SEQ ID No.: 74	Nucleotide fragment of the third coding exon strain Heteren-P-S
SEQ ID No.: 75	Nucleotide fragment of the third coding exon strain Ling-Q-S
SEQ ID No.: 76	Nucleotide fragment of the third coding exon strain Mao-Q-S
SEQ ID No.: 77	Nucleotide fragment of the third coding exon strain TemR-Q-S
SEQ ID No.: 78	Nucleotide fragment of the third coding exon strain Uppsala-T***-S
SEQ ID No.: 79	Nucleotide fragment of the third coding exon strain Trans-Q-S
SEQ ID No.: 80	Nucleotide fragment of the third coding exon strain BEQ-Q-S

SEQ ID No.: 81	Nucleotide fragment of the third coding exon strain BSQ-Q-S
SEQ ID No.: 82	Nucleotide fragment of the third coding exon strain Brazza-Q-S
SEQ ID No.: 83	Nucleotide fragment of the third coding exon strain Bouaké-Q-R
SEQ ID No.: 84	Nucleotide fragment of the third coding exon strain Thai-Q-S
SEQ ID No.: 85	Nucleotide fragment of the third coding exon strain Madurai-Q-S
SEQ ID No.: 86	Nucleotide fragment of the third coding exon strain Recife-Q-R
SEQ ID No.: 87	Nucleotide fragment of the third coding exon strain Brésil Q-S
SEQ ID No.: 88	Nucleotide fragment of the third coding exon strain Moorea Q-S
SEQ ID No.: 89	Nucleotide fragment of the third coding exon strain Killcare P-S
SEQ ID No.: 90	(1)
SEQ ID No.: 91	(1)
SEQ ID No.: 92	(1)
SEQ ID No.: 93	(1)
SEQ ID No.: 94	(1)
SEQ ID No.: 95	(1)
SEQ ID No.: 96	(1)
SEQ ID No.: 97	(1)
SEQ ID No.: 98	(1)
SEQ ID No.: 99	(1)
SEQ ID No.: 100	(1)
SEQ ID No.: 101	(1)
SEQ ID No.: 101	(1)
SEQ ID No.: 102	(1)
SEQ ID No.: 103	(1)
SEQ ID No.: 104	(1)
SEQ ID No.: 105	(1)
SEQ ID No.: 106	(1)
SEQ ID No.: 107	(1)
SEQ ID No.: 108	(1)
SEQ ID No.: 109	(1)
SEQ ID No.: 110	(1)
SEQ ID No.: 111	(1)
SEQ ID No.: 111	(1)
SEQ ID No.: 112	(1)
SEQ ID No.: 113	(1)
SEQ ID No.: 114	(1)
SEQ ID No.: 115	(1)
SEQ ID No.: 116	(1)
SEQ ID No.: 117	(1)
SEQ ID No.: 118	(1)
SEQ ID No.: 119	(1)
SEQ ID No.: 120	*An. gambiae* strain YAO *ace-1* gene
SEQ ID No.: 121	*An. gambiae* strain YAO AChE1 cDNA (complete coding sequence)
SEQ ID No.: 122	*An. gambiae* strain YAO AChE1 protein (complete sequence)
SEQ ID No.: 123	Primer Ex3 AG dir

SEQ ID No.: 124	Primer Ex3 AG rev
SEQ ID No.: 125	*An. gambiae* strain KISUMU AChE1 cDNA (complete sequence)
SEQ ID No.: 126	*An. gambiae* strain KISUMU AChE1 protein (complete sequence)
SEQ ID No.: 127	*An. gambiae ace-1* gene (including 2 5' non-coding exons)
SEQ ID No.: 128	Primer Moustdir1
SEQ ID No.: 129	Primer Moustrev1

* P = *Culex pipiens pipiens* (subspecies pipiens)
** Q = *Culex pipiens quinquefasciatus* (subspecies quinquefasciatus)
*** T = *Culex torrentium*
**** R = resistant
***** S = sensitive
(1) peptide fragments of the third coding exon corresponding to the nucleotide fragments SEQ ID No.s. 60 to 89

The nucleotide sequences (SEQ ID NOs. 27 to 38) and the corresponding peptide sequences (SEQ ID NOs. 10 to 21) were submitted to various sequence libraries on March 8, 2002, but were only made accessible on November 30, 2002, in the EMBL sequence base and on January 11, 2003, in the GENBANK sequence base.

EXAMPLE 1: Materials and methods

a) Strains and crosses

Five *C. pipiens* strains were used: S-LAB, a standard insecticide-sensitive strain (Georghiou et al., 1996, Bull. Wld. Hlth Org., 35, 691-708), SA1, SA4 and EDIT, which have a single insecticide-sensitive AChE, and SR which is homozygous for an insecticide-insensitive AChE (Berticat et al., Genet. Res., 2002, 79, 41-47). The strains having a sensitive and insensitive AChE are referred to, respectively, as S and R.

b) *ace* gene nomenclature and amino acid sequence numbering

ace-1 represents the locus encoding a cholinergic AChE responsible for resistance to organo-phosphorus compounds and/or to carbamates in *C. pipiens* (AChE1), previously called *Ace.1* (Raymond et al., Genetica, 2001, 112/113, 287-296). *ace-2* represents the second *ace* locus, which is not involved in insecticide resistance in *C. pipiens* (previously called *Ace.2*), the

function of which is unknown in *C. pipiens*. The single *ace* gene present in *Drosophila melanogaster*, which is homologous to *ace-2*, is therefore similarly named.

In the analyses which follow, the positions of the amino acid residues are indicated with reference to the sequence of the torpedo fish AChE [*Torpedo californica*; GENBANK P04058], according to the nomenclature recommended by Massoulié et al., 1992, *In Multidisciplinary approaches to cholinesterase functions, eds, Schafferman, A. & Velan, B. (Plenum Press New York), p 285-288*].

c) Analysis of transmission of the *ace-1* gene

With the females being indicated first, F1 crosses (S X R) and back crosses (F1 X S-LAB and S-LAB X F1) were obtained by mass crossing of adults. A few larvae derived from the back crosses were treated with a dose of carbamate (propoxur, 4 mg/l) which kills 100% of sensitive larvae. The linkage between *ace-1* and propoxur resistance was studied by RFLP in the surviving larvae, based on a 320 bp PCR product making it possible to identify the S and R alleles. The experiments were carried out independently, with S = SA1, S = SA4 and S = EDIT.

d) Sequence analysis and gene assembly

All the sequence analyses were carried out based on the crude sequences of *Anopheles gambiae* available on the INFOBIOGEN server (http://www.infobiogen.fr) and the tools available on the site (http://www.ncbi.nlm.nih.gov/blast/blast). The genomic sequences encoding an AChE were identified using the TBLASTN and BLAST programs (Altschul et al., J. Biol. Mol., 1990, 215, 403-410). The genomic sequences identified were assembled using the ABI Prism Auto-Assembler program (v2.1, Perkin Elmer). The sequences were verified and corrected using the Ensembl Trace Server program (http://trace.ensembl.org/). Two concatenations of, respectively, 5195 and 6975 base pairs, encoding respectively AChE1 and AChE2, were assembled from, respectively, 64 and 74 independent

sequences (mean redundancy of 10.5 and 6.5). The exons and the protein sequences were identified using a combination between the FGENESH (http://www.sanger.uk) and BLASTX (http://www.ncbi.nlm.nih.gov) programs. The genomic sequences of ascidian AChE were assembled from crude sequences deposited in the databases of the NCBI (*Ciona savignyi*) and of the Doe Joint Institute (*Ciona intestinalis*, http://www.jgi.doe.gov/programs/ciona/ciona-mainage.html). The searches in the *Drosophila* databases were carried out using Flybase (http://www.fruitfly.org/).

e) Sequence comparisons

The sequences of the *Anopheles gambiae* AChE1 and AChE2 proteins deduced from the genomic sequences and the peptide sequences deduced from PCR fragments of *C. pipiens* and *A. aegypti* were aligned with those of known AChEs, by means of the ClustalW program, using a BLOSUM matrix and default parameters (Thompson et al., N.A.R., 1994, 22, 4673-4680).

A phylogenetic tree was constructed using the neighbor-joining algorithm of version DDBJ of Clustal W (http://hypernig.nig.ac.jp/homology/ex-clustalw-e.shtml). *Bootstrap* analysis (1000 counts and 111 entry values) was used to evaluate the degrees of confidence for the topology of the tree. The construction of the trees was carried out using the Treeview program (v1.6.6).

f) Accession numbers

The numbers of the sequences (accession numbers in the databases or the identifying numbers in the sequence listing) which were used for the genetic analysis are as follows:

- *Craniata*: *Homo sapiens*: NP_00046; *Bos taurus*: P23795; *Felix catus*: O6763; *Oryctolagus cuniculus*: Q29499; *Rattus norvegicus*: P36136; *Mus musculus*: P21836; *Gallus gallus*: CAC37792; *Danio reno*: Q9DDE3; *Electrophorus electricus*: 6730113; *Torpedo marmorata*: P07692; *Torpedo californica*: P04058; *Bungarus fasciatus*: Q92035; *Myxine glutinosa*: Q92081.

- Cephalocordes: *Branchiostoma floridae*: O76998 and 076999; *Branchiostoma lanceolatum*: Q95000 and Q95001.

- Urocordes: *Ciona intestinalis*: SEQ ID NO 51; *Ciona savignyi*: SEQ ID NO 52.

- Nematodes: *Caenorhabditis elegans* (1 to 4): P38433, O61371, O61459 and 061372; *Caenorhabditis briggsae* (1 to 4) Q27459, O61378Q9NDG9 and Q9NDG8; *Dictyocaulus viviparus*: Q9GPLO.

- Insects: *Anopheles gambiae* (1 and 2): SEQ ID NO 3 and SEQ ID NO 53 (BM 629847 anD BM 627478); *Aedes aegypti* (1 and 2): SEQ ID NO 9 and AAB3500; *An. stephensi*: P56161; *Culex pipiens*: SEQ ID NO 7 (*ace-1*) and Esther database for *ace-2*; *Drosophila melanogaster*: P07140; *Lucilia cuprina*: P91954; *Musca domestica*: AAK69132.1; *Leptinotarsa decemlineata*: Q27677; *Apis mellifera*: AAG43568; *Nephotettix cincticeps*: AF145235_1; *Schizaphis graminum*: Q9BMJ1.

- Arachnids: *Rhipicephalus appendiculatus*: O62563; *Boophilus microplus* (1 and 2): O45210 and 061864; *Boophilus decoloratus*: O61987.

- Mollusks: *Loligo opalescens*: O97110.

g) Cloning of the fragment K and genotyping of *ace-1* in *Culex pipiens*

The mosquito DNA was extracted as described in Rogers et al. [*Plant Molecular Biology manual, 1988, eds. Gelvin, S.B.1 Schilperoot, R.A. (Kluwer Academic Publishers, Boston), Vol. A6, p 1-10*]. The oligonucleotides PkdirAGSG (5'-ATMGWGTTYGAGTACACSGAYTGG-3', SEQ ID NO 39) and Pkrev AGSG (5'-GGCAAARTTKGWCCAGTATCKCAT-3', SEQ ID NO 40) amplify a 320 bp fragment (fragment K) from the genomic DNA of several mosquitoes. 30 PCR amplification cycles were carried out under the following conditions: 94°C for 30 s, 50°C for 30 s and 72°C for 30 s. The sequences were determined directly on the PCR products on an ABI prism 310 sequencer, using the *Big Dye Terminator* kit.

The genotyping of *ace-1 Culex* was carried out under the following conditions: the fragments K obtained

as described above were digested with *EcoRI* and the digestion product was separated by electrophoresis on a 2% agarose gel. The restriction profiles show: 1 band (320 bp) in the resistant homozygous RR mosquitoes, 2 bands (106 bp and 214 bp) in the homozygous SS mosquitoes and 3 bands (103 bp, 214 bp and 320 bp) in the heterozygous RS mosquitoes.

h) Cloning of the *ace-1* cDNA in sensitive and resistant individuals

The cDNA of the *Culex pipiens ace-1* gene was obtained from the RNA extracted from individuals of the reference sensitive strain S-LAB and of the resistant strain SR, at the very first larval stage of the development L1. The reverse transcription was carried out with an 18T oligonucleotide and SuperScriptIIRNaseH (INVITROGEN), according to the conditions recommended by the manufacturer.

- Strain S-LAB

Two cDNA fragments were amplified by PCR using degenerate oligonucleotides obtained from the alignment of the sequences of the *Anopheles gambiae* and *Schizaphis graminum ace-1* genes:

- a fragment b (193 bp) was amplified using the pair of primers PbdirAGSG (5'GGYGCKACMATGTGGAAYCC3', SEQ ID NO 41) and PbrevAGSG (5'ACCAMRATCACGTTYTCYTCCGAC3', SEQ ID NO 42);
- a fragment k (320 bp) was amplified using the pair of primers PkdirAGSG (5'ATMGWGTTYGAGTACACSGAYTGG3', SEQ ID NO 39) and PkrevAGSG (5'GGCAAARTTKGWCCAGTATCKCAT3', SEQ ID NO 40).

The fragments b and k thus obtained were then cloned and sequenced, according to conventional techniques known in themselves to those skilled in the art, as described in Current Protocols in Molecular Biology (*Frederick M. AUSUBEL, 2000, Wiley and Son Inc, Library of Congress, USA*).

A larger cDNA fragment was amplified by PCR, using the *Culex pipiens*-specific primers deduced from

the sequences of the fragments b and k obtained above. Namely:

- a fragment CulexA (1127 bp) was amplified by PCR using the pair of primers: culex-bdir1 (5'TACATCAACGTGGTCGTGCCACG3', SEQ ID NO 43) and culex-krev1 (5'GTCACGGTTGCTGTTCGGG3', SEQ ID NO 44). The 1127 bp fragment CulexA thus obtained was then cloned and sequenced, as above.

The ends of the cDNAs were amplified by the RACE (*Rapid Amplification of cDNA Ends*) technique using a commercial kit (*Gene Racer* kit (INVITROGEN)) according to the conditions indicated in the instruction booklet. They were subsequently cloned and then sequenced, as above.

- Strain SR

The complete sequence of the cDNA of the *ace-1* gene of the resistant strain SR was amplified by PCR using the primers culex-5'dir (5'-CCACACGCCAGAAGAAAAGA-3', SEQ ID NO 54) and culex-3'dir (5'-AAAAACGGGAACGGGAAAG-3', SEQ ID NO 55) and the 2497 bp fragment thus obtained was cloned and sequenced, as above.

i) Cloning of the *ace-1* gene in sensitive and resistant individuals

The genomic DNA of the strain KISUMU (reference sensitive strain from West Africa) and of the strain YAO (resistant strain isolated in Ivory Coast) of *A. gambiae* was extracted from homozygous individuals as described in Rogers et al. [*Plant Molecular Biology manual, 1988, eds. Gelvin, S.I.1 Schilperoot, R.A. (Kluwer Academic Publishers, Boston) Vol. A6, p 1-10*].

3 overlapping fragments (A, B and C) were amplified under the following conditions: 94°C for 30 s, 50°C for 30 s and 72°C for 30 s (30 cycles), using the primers synthesized from the sequence of the *ace-1* gene. Namely:

- the fragment A (1130 bp) was amplified using the pair of primers AG1-Adir (5'CGACGCCACCTTCACA3', SEQ ID NO 45) and AG1-Arev (5'GATGGCCCGCTGGAACAGAT3', SEQ ID NO 46),

- the fragment B (1167 bp) was amplified using the pair of primers AG1-Bdir (5'GGGTGCGGGACAACATTCAC3', SEQ ID NO 47) and AG1-brev (5'CCCCGACCGACGAAGGA3', SEQ ID NO 48), and

- the fragment C (876 bp) was amplified using the pair of primers AG1-Cdir (5'AGATGGTGGGCGACTATCAC3' SEQ ID NO 49) and AG1-Crev (5'CTCGTCCGCCACCACTTGTT3', SEQ ID NO 50).

The sequences of the fragments A, B and C were determined directly on the PCR products, by means of internal oligonucleotides, included in these fragments, using the *Big Dye Terminator* kit and an ABI prism 310 sequencer.

j) Detection of the mutation of the third coding exon responsible for the insecticide resistance in mosquitoes of the species *C. pipiens* and *An. gambiae*

The mosquito DNA was extracted as described in Rogers et al., mentioned above, and a fragment of the third coding exon was then amplified by PCR and sequenced, and the mutation in the coding sequence of the third coding exon was detected by PCR-RFLP, according to the principle as described above for the fragment K.

- *C. pipiens*

A 520 bp fragment of the third coding exon was amplified from the genomic DNA of several mosquitoes, by PCR using the pair of primers:

- Ex3dir 5'-CGACTCGGACCCACTGGT-3' (SEQ ID NO 58) and
- Ex3rev 5'-GTTCTGATCAAACAGCCCCGC-3' (SEQ ID NO 59).

The fragment thus obtained was digested with *Alu I* and the digestion product was separated by electrophoresis on a 2% agarose gel. The expected restriction profiles are as follows: 1 fragment (520 bp) in the sensitive homozygous SS individuals, 2 fragments (357 bp and 163 bp) in the resistant homozygous RR individuals and 3 fragments (520 bp, 357 bp and 163 bp) in the resistant heterozygous RS individuals.

- *An. gambiae*

A 541 bp fragment of the third coding exon was amplified from the genomic DNA of several individuals, by PCR using the pair of primers:

- Ex3AGdir 5'-GATCGTGGACACCGTGTTCG-3' (SEQ ID NO 123) and
- Ex3AGrev 5'-AGGATGGCCCGCTGGAACAG-3' (SEQ ID NO 124).

The fragment thus obtained was digested with *Alu I* and the digestion product was separated by electrophoresis on a 2% agarose gel. The expected restriction profiles are as follows: 2 fragments (403 bp and 138 bp) in the sensitive homozygous SS individuals, 3 fragments (253 bp, 150 bp and 138 bp) in the resistant homozygous RR individuals and 4 fragments (403 bp, 253 bp, 150 bp and 138 bp) in the resistant heterozygous RS individuals; given that the 150 bp and 138 bp fragments comigrate, the resistant homozygous and heterozygous individuals are detected, respectively, by the presence of 2 bands (253 bp and approximately 150 bp) and of 3 bands (403 bp, 253 bp and approximately 150 bp) in agarose gel.

- *C. pipiens, An. gambiae* and *An. albimanus*

A 174 bp fragment of the third coding exon was amplified from the genomic DNA of several mosquitoes, by PCR using the pair of primers:

- Moustdir1: 5' CCGGGNGCSACYATGTGGAA 3' (SEQ ID NO 128), and
- Moustrev1: 5' ACGATMACGTTCTCYTCCGA 3'(SEQ ID NO 129).

The fragment thus obtained was digested with *Alu I* and the digestion product was separated by electrophoresis on a 2% agarose gel. The expected restriction profiles are as follows: 1 fragment (194 bp) in the sensitive homozygous SS individuals, 2 fragments (74 bp and 120 bp) in the resistant homozygous RR individuals and 3 fragments (194 bp, 74 bp and 120 bp) in the resistant heterozygous RS individuals.

The results are illustrated in figure 11.

Figure 12 shows that, with resistant *An. albimanus* mosquitoes, the same propoxur-inhibition characteristics are obtained, by means of a conventional biochemical test, as for the *An. gambiae* and *C. pipiens* mosquitoes.

Application of the diagnostic test, also referred to as "G119S", using the Moustdir1 and Moustrev1 amplimer reveals the presence of an *AluI* site associated with resistance (figure 11). Sequencing of the amplified fragments of *An. albimanus* confirms the substitution of the Gly codon GGC, in the SS individuals, to an Ser codon AGC in the RR individuals (figures 13 and 14).

k) Measurement of the acetylcholinesterase activity

The cDNAs encoding the AChE1s of, respectively, the strain S-LAB and the strain SR were cloned in the eukaryotic expression vector pAc5.1/V5-His (INVITROGEN), according to conventional recombinant DNA techniques, by following standard protocols such as those described in *Current Protocols in Molecular Biology*, mentioned above. Drosophila cells (Schneider S2 cells) were transfected with the recombinant vectors thus obtained, using the Fugen® reagent (ROCHE), according to the manufacturer's instructions. 24 hours after transfection, the cells were harvested by centrifugation and then lyzed in a 0.25M phosphate buffer containing 1% Triton X-100. The acetylcholinesterase activity of the cell extracts obtained was measured, in the presence or in the absence of insecticide (propoxur), by the method as described in Bourguet et al., Biochemical Genetics, 1996, 34, 351-362.

EXAMPLE 2: Demonstration of 2 *ace* genes in *Anopheles gambiae*

Genes homologous to the human and drosophila acetylcholinesterase genes were sought based on sequence fragments deposited in the databases, using the TBLASTN program. Two groups of distinct fragments encoding an AChE very similar to that of drosophila were identified. Two genes of, respectively, 6975 bp

(*ace-1*) and 5195 bp (*ace-2*) were reconstructed from overlapping fragments of each group. Analysis of the genes using the FGENESH and BLASTX programs shows that the *ace-1* and *ace-2* genes consist, respectively, of at least 7 and 8 exons encoding proteins of approximately 534 and 569 amino acids, respectively referred to as AChE1 and AChE2. However, this analysis did not make it possible to determine with certainty the sequence of the 5′ and 3′ ends of the cDNA and the NH_2 and COOH sequences of the corresponding proteins, which are not conserved between the various AChEs.

The amino acid sequence analysis confirms that the AChE1 and AChE2 proteins are highly homologous to the AChE of *Drosophila* (BLASTP : $P < e^{-180}$) and contain a canonic FGESAG motif around the serine at position 200, with reference to the sequence of the *Torpedo* AChE (S_{200}, figure 1), which is characteristic of the AChE active site. In addition, other motifs characteristic of AChEs were also found in the two sequences (AChE1 and AChE2): the choline-binding site (tryptophan residue at position 84, W84), the three residues of the catalytic triad (serine, glutamic acid and histidine residues, respectively at positions 200, 327 and 440: S_{200}, E_{327} and H_{440}), the six cysteine residues potentially involved in conserved disulfide bridges (C_{67}-C_{94}; C_{254}-C_{265}; C_{402}-C_{521}), and aromatic residues bordering the active site gorge (10 and 11 residues, respectively, for AChE1 and AChE2).

In the two sequences, the presence of a phenylalanine residue is observed at position 290 (F290), but not at position 288; this characteristic common to invertebrate AChEs is responsible for a broader substrate specificity of invertebrate AChEs, compared with those of vertebrates.

Analysis of the C-terminal sequences of Diptera AChEs shows the presence of a hydrophobic peptide corresponding to a signal for the addition of a glycolipid, indicating post-translational cleavage of a C-terminal fragment and the addition of a glycolipid

anchoring residue as in *Drosophila*, and other mosquito species. In all the sequences, the presence of a cysteine residue is also observed in the C-terminal sequence preceding the potential site of cleavage of the hydrophobic peptide. This cysteine residue could be involved in an intermolecular disulfide bond, linking the two catalytic subunits of the AChE dimer.

The AChE1 and AChE2 proteins of *An. gambiae* exhibit 53% similarity between one another and show, respectively: 76% and 55% similarity with the AChE of *Schizaphis graminum* (NCBI accession number AAK09373 or GENBANK accession number 12958609), 53% and 98% similarity with the AChE of *An. stephensi* (GENBANK 2494391), 54% and 95% similarity with the AChE of *Aedes aegypti* (GENBANK 2133626), and 52% and 83% similarity with the AChE of *Drosophila* (GENBANK 17136862).

The major difference between AChE1 and AChE2 lies in an insertion of 31 amino acids in the sequence of AChE2 (figure 1). This sequence, referred to as "hydrophilic insertion" in the AChE of *Drosophila* is absent in vertebrate and nematode AChEs and could be characteristic of AChE2, at least in the diptera.

These results demonstrate the presence of two *ace* genes in the genome of *Anopheles gambiae*, one encoding AChE1 which is related to the AChE of *Schizaphis graminum*, and the other encoding AChE2 which is related to the AChE of *Drosophila* and to the known AChEs of mosquitoes. The presence of other *ace* genes in *An. gambiae* is highly improbable insofar as complementary searches in the databases of the *An. gambiae* genome, using less stringent parameters, detected only sequences encoding alpha-esterases (EC 3.1.1) and carboxylesterases (EC 3.1.1.1).

EXAMPLE 3: Demonstration of a single *ace* gene in *Drosophila melanogaster*

The presence of a gene homologous to the *ace-1* gene was sought in the genome of *Drosophila*. TBLASTN searches made it possible to detect the *ace* gene identified above, homologous to the *ace-2* gene of

Anopheles gambiae, but did not make it possible to detect other sequences homologous to the *ace-1* gene. Searches using less stringent parameters made it possible to detect only alpha- and carboxylesterases. These results demonstrate that the drosophila genome contains a single *ace* gene (*ace-2*).

EXAMPLE 4: Demonstration of at least two *ace* genes in the other mosquito species

The presence of the *ace-1* gene in the genome of other mosquito species was analyzed by PCR using degenerate oligonucleotides (PdirAGSG and PrevAGSG, SEQ ID NOs. 39 and 40) for amplifying an exon fragment (fragment K of approximately 320 bp, figure 1), corresponding to sequences that are conserved between the AChE1 sequences of *An. gambiae* and *Schizaphis graminum*, but divergent between the AChE1 and AChE2 sequences of *An. gambiae*.

The sequence of the PCR products obtained from the genomic DNA of various mosquito species shows a very high percentage identity between the sequences of *Anopheles*, *Culex* and *Aedes*. In addition, most of the substitutions are silent since the amino acid sequences deduced from these nucleotide sequences only differ from one another by 5 to 6 amino acids (figure 2A). The fragment K was also amplified by RT-PCR from the mRNA of *C. pipiens*, indicating that the *ace-1* gene is expressed in the form of mRNA; this result is in agreement with the existence, in *C. pipiens*, of two AChEs having distinct catalytic properties.

EXAMPLE 5: Analysis of the linkage between the *ace-1* gene and insecticide resistance

In order to analyze the linkage between the *ace-1* gene and insecticide resistance, the fragment K amplified from the genomic DNA of resistant *C. pipiens* (strain R) was sequenced. Comparison of the fragment K sequences between the S and R strains shows differences of 3 nucleotides (silent substitutions, figure 2B). One of these substitutions affects an *EcoRI* site, which makes it possible to readily differentiate the *ace-1*

loci of the S and R strains by PCR-RFLP: the restriction profiles show 1 band (320 bp) in the resistant homozygous individuals, 2 bands (106 bp and 214 bp) in the homozygous SS mosquitoes and 3 bands (103 bp, 214 bp and 320 bp) in the heterozygous RS mosquitoes (figure 2C).

The linkage between the *ace-1* gene and propoxur resistance was studied, in triplicate, in the following way: larvae from a backcross (S x R) x S were treated for a dose that is lethal for sensitive individuals and the *ace-1* genotype was analyzed in the survivors, by PCR-RFLP.

The results show that exposure to propoxur kills 50% of the larvae in all the back crosses, i.e. all the sensitive individuals. All the surviving larvae (100 for each back cross, 300 in total) show a heterozygous profile by RFLP, indicating that they all have a copy of the *ace-1* gene of the R strain.

These results demonstrate that the *ace-1* gene is very closely linked with the insecticide resistance (less than 1% recombination with a degree of confidence of 0.05).

EXAMPLE 6: Analysis of the phylogeny of the *ace-1* and *ace-2* genes

Phylogenetic trees were constructed from the sequences of the conserved regions of *An. gambiae* AChEs (SEQ ID NO 1 and fragment 34-393 of the sequence SEQ ID NO 53, figure 1), of the fragments K of *C. pipiens* and *Aedes aegypti* (SEQ ID NOs. 8 and 9) and from 33 AChE sequences available in GENBANK, using the neighbor-joining method, as described in the materials and methods.

Figure 3 illustrates the heterogeneity of the number of *ace* genes in the course of the evolution of the animal kingdom. Among the chordates, the cephalochordates have at least two *ace* genes, whereas the urochordates have only a single one, as deduced from the analysis of their genome. Among the arthropods, the Diptera have either a single *ace* gene (*Drosophila* of

the suborder Brachycera) or two *ace* genes (mosquitoes of the suborder nematocera). The topology of the tree shows that these two *ace* genes became duplicated very early in the course of evolution, probably before separation between protostomes and deuterostomes. These results are supported by the fact that mollusk, nematode and arthropod AChEs branch from the sequences of chordata (*craniatia*, cephalochordata and urochordata). The results show that arthropods and nematodes have a related AChE.

These results indicate that the *ace-1* and *ace-2* genes identified in insects originate from a very distant duplication event and that the absence of the *ace-1* gene, at least in certain species of the suborder Brachycera (*Drosophila*), results from the loss of an *ace* gene rather than from a recent duplication of the *ace* gene in the nematocera. These results also suggest that the extrapolations made from studies in *D. melanogaster* are to be considered with reservation insofar as the situation in *Drosophila* is representative neither of the Diptera nor of the entire insect class.

EXAMPLE 7: Determination of the cDNA sequence of the *ace-1* gene

The *ace-1* cDNA was cloned from two strains of *Anopheles gambiae* (sensitive strain KISUMU and resistant strain YAO) and from two strains of *Culex pipiens* (sensitive strain S-LAB and resistant strain SR), as described in the materials and methods.

The complete sequence of the cDNA of the KISUMU strain corresponds to the sequence SEQ ID NO 125 which encodes a 737 amino acid protein (SEQ ID NO 126). The complete sequence of the cDNA and of the AChE1 protein of the strain YAO correspond, respectively, to the sequences SEQ ID NO 121 and SEQ ID NO 122.

The sequences SEQ ID NO 4 and SEQ ID NO 5 correspond to the virtually complete sequence (with the exception of the first coding exon of the *ace-1* gene),

respectively, of the cDNA and of the AChE1 protein of the strain KISUMU.

The complete sequence of the cDNA of the *C. pipiens* strains S-LAB and SR correspond, respectively, to the sequences SEQ ID NO 6 and SEQ ID NO 56 which encode a 702 amino acid protein (SEQ ID NO 7 and SEQ ID NO 57, respectively, for the strain S-LAB and the strain SR).

EXAMPLE 8: Determination of the sequence of the *ace-1* gene

The sequence of the *ace-1* gene was determined from the genomic DNA of two strains of *Anopheles gambiae*, the reference sensitive strain from West Africa (strain KISUMU) and a resistant strain from Ivory Coast (strain YAO), as described in the materials and methods.

The complete *An. gambiae* sequence corresponds to the sequence SEQ ID NO 127 which has an intron-exon organization comprising at least 9 exons and including two 5' non-coding exons (table I).

The virtually complete sequence (with the exception of the first two 5' non-coding exons) of the *ace-1* gene of the strain KISUMU corresponds to the sequence SEQ ID NO 23.

The virtually complete sequence (with the exception of the first two 5' non-coding exons and of the first coding exon) of the *ace-1* gene of the strain YAO corresponds to the sequence SEQ ID NO 120.

EXAMPLE 9: Identification of (a) mutation(s) in the amino acid sequence of the AChE1 protein that is(are) responsible for the insecticide resistance in mosquitoes of the species *Culex pipiens* and *Anopheles gambiae*

The nucleotide sequence encoding the AChE1 protein (cDNA) was determined from two strains of *Anopheles gambiae* (sensitive strain KISUMU and resistant strain YAO) and from two strains of *Culex pipiens* (sensitive strain S-LAB and resistant strain SR), as described in example 7.

The amino acid sequences of the AChE1 protein of the sensitive and resistant strains, deduced from the above sequences, were then aligned (figures 5, 6 and 9).

Comparison of the amino acid sequences of the *C. pipiens* AChE1 protein (figures 5 and 6) shows that a single non-silent mutation exists between the insecticide-sensitive strain (S-LAB, SEQ ID NO 7) and the insecticide-resistant strain (strain SR, SEQ ID NO 57), located in the region encoded by the third coding exon of the *ace-1* gene: the glycine (GGC) at position 247 (or at position 119, with reference to the sequence of the torpedo fish AChE) of the sensitive strain is replaced with a serine (AGC) in the resistant strain ($G_{247(119)} \rightarrow S_{247(119)}$).

The location of the amino acid at position 247 in the *C. pipiens* acetylcholinesterase structure and the effect of the glycine → serine substitution on this structure were analyzed by molecular modelling based on the structure of the torpedo fish acetylcholinesterase. The results are given in figure 7 (A, B and C). Figure 7A shows that the amino acid at position 119 is close to the residues of the catalytic site (S_{200} and H_{440}). Figure 7C shows that, by comparison with the glycine of the sensitive strain (figure 7B), the hindrance of the side chain of the serine in the resistant strain reduces the space of the catalytic site, which probably prevents the insecticide from interacting with the catalytic serine (S_{200}).

Comparison of the amino acid sequences of the *An. gambiae* AChE1 protein shows that two non-silent mutations exist between the insecticide-sensitive strain (KISUMU, SEQ ID NO 5) and the insecticide-resistant strain (strain YAO, SEQ ID NO 122): the first corresponds to the replacement of the valine (CGT) at position 33 of the sequence of the sensitive strain (SEQ ID NO 5) with an alanine (CGC) in the resistant strain, and the second is the same glycine → serine mutation as that found in *Culex pipiens*.

Given the external position of the valine in the acetylcholinesterase structure, this mutation is certainly not involved in the resistance in *Anopheles gambiae* and only the serine should be responsible for the insecticide resistance both in *Anopheles gambiae* and *Culex pipiens*.

EXAMPLE 10: Detection of the mutation in the third coding exon of the *ace-1* gene that is responsible for the insecticide resistance in mosquitoes of the species *Culex pipiens* and *Anopheles gambiae*

The restriction profile of the third coding exon of the *ace-1* gene containing the glycine → serine mutation was verified in many populations and strains of mosquitoes of the species *C. pipiens* and *An. gambiae*, sensitive and resistant to insecticides of the organophosphorus compound and carbamate class, by PCR-RFLP according to the protocol as described in example 1.

More precisely:

- in *C. pipiens*, the glycine (GGC) → serine (AGC) mutation introduces a unique *Alu I* site (AGCT) into the sequence of the resistant strain, which is demonstrated on the basis of a 520 bp PCR product amplified using the primers Ex3dir and Ex3rev, as illustrated in figure 6;

- in *An. gambiae*, the glycine (GGC) → serine (AGC) mutation introduces a second *Alu I* site (AGCT) into the sequence of the resistant strain, which is demonstrated on the basis of a 541 bp PCR product amplified using the primers Ex3AGdir and Ex3AGrev, as illustrated in figure 9.

The PCR-RFLP results were then verified by sequencing the 520 bp or 541 bp PCR fragment of the third coding exon.

- Species *C. pipiens*

The resistant (R) or sensitive (S) *Culex pipiens* strains and populations which were analyzed are given in table III below:

Table III: Strains and populations of the species *C. pipiens* analyzed

Classification	Name	R/S*	Country	Reference
C. p. quinque fasciatus	BO	R	Burkina-Faso	Isolated by the inventors
	HARARE	R	Zimbabwe	Isolated by the inventors
	SUPERCAR	R	Ivory Coast	(F. Chandre, Doctoral Thesis, University Paris XII, 1998).
	DJI	R	Mali	Isolated by the inventors
	MARTINIQUE	R	Martinique	Bourguet et al., Biochem. Genet., 1996, 34, 351-362
	RECIFE	R	Brazil	Isolated in 1995 by A.-B. Failloux, Pasteur Institute, Paris (France)
	PRO-R	S	United States	Georghiou et al., Bull. Wld Hlth Org., 1996, 35, 691-708.
	S-LAB	S	United States	Georghiou et al., Bull. Wld Hlth Org., 1996, 35, 691-708.
	TEM-R	S	United States	Georghiou et al., J. Econ. Entomol., 1978, 71, 201-205.
	TRANS-P	S	United States	Priester et al., J. Econ. Entomol., 1978, 71, 197-200.
	LING	S	China	Weill et al., J. American Mosquito Control Assoc., 2001, 17, 238-244
	THAI	S	Thailand	Guillemaud et al., Heredity, 1996, 77, 535-543.
	MAO	S	China	Qiao et al., Biochem. Genet., 1998, 36, 417-426.
	MADURAI	S	India	Nielsen-Leroux, et al., J. Med. Entomol., 2002, 39, 729-735
	BSQ	S	South Africa	Isolated in 1991 by A.J. Cornel (Sth Afr. Inst. Med. Res., South Africa)
	BED	S	South Africa	Isolated in 1991 by A.J. Cornel (Sth Afr. Inst. Med. Res., South Africa)
	BOUAKE	S	Ivory Coast	Magnin et al., J. Med. Entomol., 1988, 25, 99-104
	BRAZZA	S	Congo	Beyssat-Arnaouty, Doctoral Thesis, University of Montpellier II (1989).
	BRESIL	S	Brazil	Isolated by the inventors
	MOOREA	S	Polynesia	N. Pasteur, et al., Genet. Res., 1995, 66, 139-146
C. p. pipiens	ESPRO	R	Tunisia	H. Ben Cheikh et al., J. Am. Mosquito Control Assoc., 1993, 9, 335-337
	PRAIAS	R	Portugal	Bourguet et al., J. Econ. Entomol., 1996, 89, 1060-1066
	PADOVA	R	Italy	Bourguet et al., Genetics, 1997, 147, 1225-1234.
	BARRIOL	R	France	Chevillon et al., Evolution, 1995, 49, 997-1007.
	BRUGES-A	S	Belgium	Raymond et al., Genet. Res., 1996, 67, 19-26.
	BRUGES-B	S	Belgium	Raymond et al., Genet. Res., 1996, 67, 19-26.
	KILLCARE	S	Australia	Guillemaud et al., Proc. R. Soc. Lond. B, 1997, 264, 245-251.
	BLEUET	S	France	Rioux et al., C.R. Séances Soc. Biol. Fil., 1961, 155, 343-344
	HETEREN	S	The Netherlands	Isolated by the inventors
C. torrentium	UPPSALA	S	Sweden	M. Raymond, Ent. Tidskr., 1995, 116, 65-66.

* R/S resistant or sensitive to insecticides of the organophosphorus compound and carbamate class

The PCR-RFLP analysis of all the mosquitoes of table III shows that a perfect correlation exists between the insecticide resistance and the restriction profile by PCR-RFLP, namely: 1 band (520 bp) is detected in the sensitive homozygous SS individuals, 2 bands (357 bp and 163 bp) are detected in the resistant homozygous RR individuals and 3 bands (520 bp, 357 bp and 163 bp) are detected in the resistant heterozygous RS individuals (figure 8).

These results were confirmed by sequencing the 520 bp PCR product for all the mosquitoes of table III analyzed by PCR-RFLP. Alignment of the sequences obtained (SEQ ID NOs. 60 to 89), illustrated in table IV below, shows that in the mosquitoes of the species *C. pipiens*, the glycine → serine mutation, located at position 739 with reference to the cDNA sequence of the *ace-1* gene of the reference sensitive strain (strain S-LAB), which is responsible for the insecticide resistance, originates from two groups of independent mutations, respectively, in *C. pipiens pipiens* and *C. pipiens quinquefasciatus*.

Table IV: Analysis of the origin of the glycine → serine mutation responsible for the insecticide resistance in mosquitoes of the species *C. pipiens*

Strains of *C. pipiens*	Position of the mutations with reference to the cDNA sequence of the *ace-1* gene of the strain S-LAB (SEQ ID NO. 6)																													
	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7	8	8	
	5	5	7	9	1	2	6	7	9	0	5	6	8	8	9	9	1	3	3	4	4	5	6	7	7	8	9	9	1	4
	0	3	1	8	3	8	4	3	7	3	1	0	1	4	1	6	4	2	9	1	7	6	3	4	7	0	0	8	3	6
C. pipiens quinquefasciatus																														
BO(R)*	T	C	A	T	C	G	G	G	G	C	G	G	G	C	C	C	C	C	A	C	C	T	C	C	C	C	G	G	A	T
Harare (R)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Supercar (R)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DJI (R)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Martinique (R)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Recife (R)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ProR(S)*	-	T	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G
S-Lab (S)	-	T	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G
TemR (S)	-	T	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	A	-	-	-
Trans (S)	-	T	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	A	-	-	
Ling (S	-	T	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	T	-	-	-	-	-	-
Thai (S)	-	T	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G
Mao (S)	-	T	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-
Madurai (S)	-	T	-	C	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G
BSQ (S)	-	T	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G
BE (S)	-	T	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-
Boualse (S)	-	T	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-
Brazza (S)	-	T	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G
Bresil(S)	-	T	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	T	-	-	-	-	-	-	-	-	G
Moorea (S)	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G
C.pipiens pipiens																														
Espro (R)	A	T	-	C	-	-	-	A	-	-	C	-	-	-	A	G	T	T	A	-	-	-	T	-	T	T	-	-	G	-
Praias (R)	A	T	-	C	-	-	-	A	-	-	C	-	-	-	A	G	T	T	A	-	-	-	T	-	T	T	-	-	G	-
Padova (R)	A	T	-	C	-	-	-	A	-	-	C	-	-	-	A	G	T	T	A	-	-	-	T	-	T	T	-	-	G	-
Barriol (R)	A	T	-	C	-	-	-	A	-	-	C	-	-	-	A	G	T	T	A	-	-	-	T	-	T	T	-	-	G	-
BrugeA (S)	A	T	-	C	-	-	-	A	-	-	C	-	-	-	A	G	T	T	G	-	-	-	T	-	T	T	-	-	G	-
BrugesB (S)	A	T	-	C	-	-	-	A	-	-	C	-	-	-	A	G	T	T	G	-	-	-	T	-	T	T	-	-	G	-
Killcare (S)	A	T	-	C	-	-	-	A	-	A	C	-	-	-	A	G	T	T	G	-	-	-	T	-	T	T	-	-	G	-
Bleuet (S)	A	T	-	C	-	-	-	A	-	-	C	-	-	-	A	G	T	T	G	-	-	-	T	-	T	T	-	-	G	-
Heteren (S)	A	T	-	C	-	A	-	A	-	A	C	-	-	-	A	G	T	-	G	-	-	-	T	-	-	T	-	A	G	-

*(R) resistant to insecticides
*(S) sensitive to insecticides

- *An. gambiae*

Sensitive strains KISUMU (reference sensitive strain from East Africa) and VK-PER (KDR reference strain from West Africa) and also sensitive populations from the Yaoundé region were tested by means of the PCR-RFLP test as described above.

The results of the PCR-RFLP test show that, for all the *An. gambiae* mosquitoes analyzed, a perfect correlation exists between the insecticide resistance and the restriction profile by PCR-RFLP, namely: 2 bands

(403 bp and 138 bp) are detected in the sensitive homozygous SS individuals, 2 bands (253 bp and approximately 150 bp) or 3 bands (403 bp, 253 bp and approximately 150 bp) are detected in the resistant individuals, respectively in the homozygous (RR) and heterozygous (RS) individuals.

EXAMPLE 11: Analysis of the acetylcholinesterase activity of the insecticide-sensitive and insecticide-resistant AChE1 proteins

The recombinant AChE1s of, respectively, the strain S-LAB and the strain SR were expressed in insect cells and the acetylcholinesterase activity was measured using cell extracts as described in example 1.

The results given in figure 10 show that the single $glycine_{247(119)} \rightarrow serine_{247(119)}$ mutation renders the acetylcholinesterase insensitive to the insecticide.

As emerges from the above, the invention is in no way limited to its methods of implementation, execution and application which have just been described more explicitly; on the contrary, it encompasses all the variants thereof which may occur to those skilled in the art, without departing from the context or the scope of the present invention.